
COFFEE LEAF DISEASE
IMAGE CLASSIFICATION
PRESERVING KENYA'S COFFEE LEGACY

Data Dragons

Victorine Imbuhila
ML Engineer

Crystal Wanjiru
Team Leader

Randell Mwania
ML Engineer

Ian Kedeyie
Data Analyst

Simon Ng’ethe
Data Analyst

Sadi Kiri
Data Engineer

Background Information

Kenya, known as the second largest African producer of Arabica coffee, boasts a rich tradition of coffee cultivation in its highland regions. However, this vital industry faces a significant threat from three major diseases—coffee rust, cercospora, and phoma—that attack coffee leaves and can devastate crops. These diseases threaten the livelihoods of small-scale farmers.

Challenges

- Disease Prevalence
- Lack of early detection
- Limited Preventative measures

Problem Statement

Kenya's highland coffee regions, renowned for their coffee cultivation tradition, are under threat from coffee rust, cercospora, and phoma diseases.

The lack of early detection methods hampers farmers' ability to prevent crop losses, posing risks to both their livelihoods and the region's coffee industry sustainability.

Data Understanding

The dataset contains leaf images collected from Arabica coffee plantation using a camera and with the help of a plant pathologist.

The data contains 29,673 images from four different classes:

- 8587 images of healthy coffee leaves
- 6571 images of Phoma
- 6200 images of Cercospora
- 8315 images of Leafrust

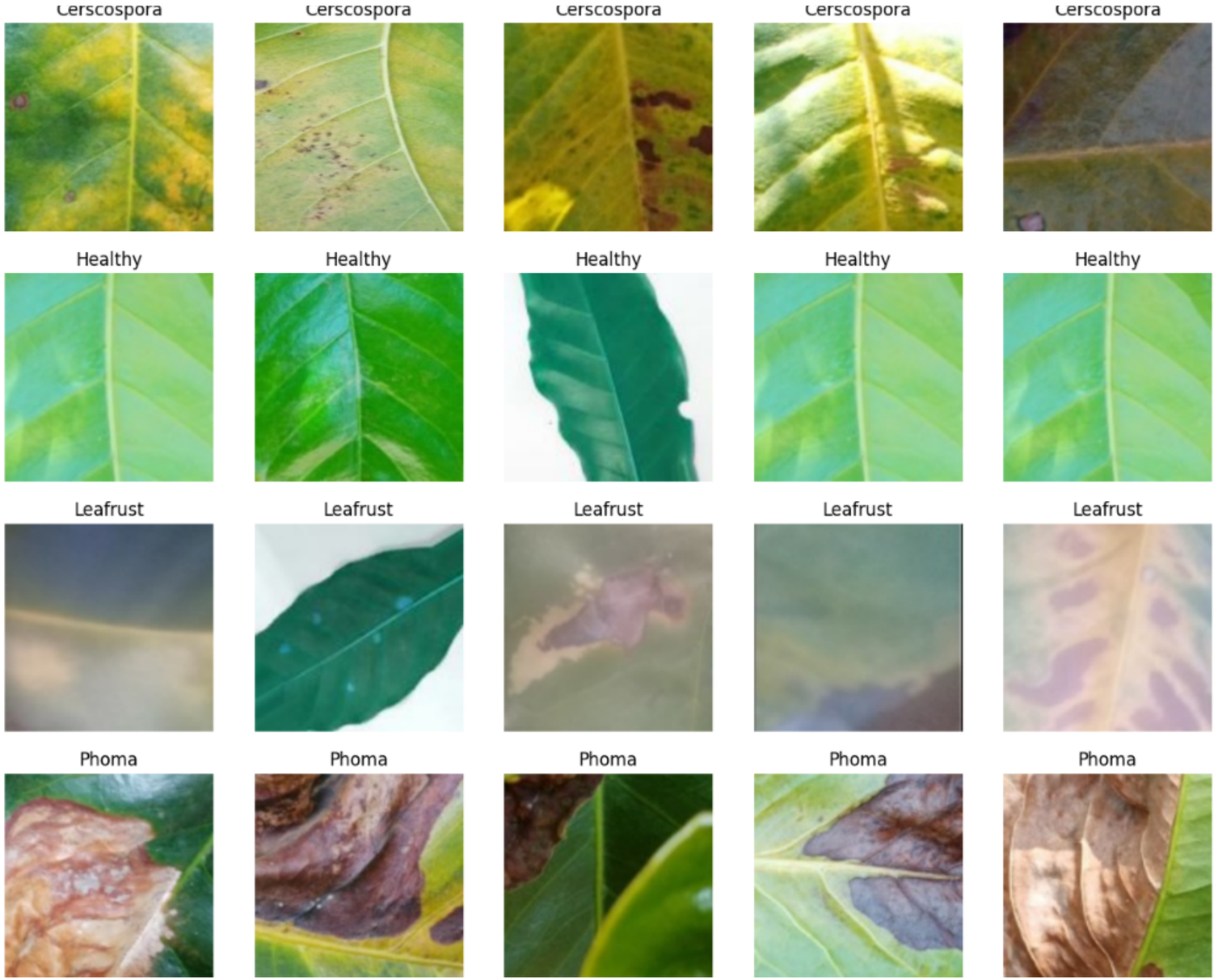
Main Objective

To develop a Convolutional Neural Network(CNN) based image classification model capable of accurately distinguishing between the three classes of coffee leaf diseases.

Specific Objectives

- To identify any patterns or anomalies in the data that may influence model performance.
- To train the model using the provided dataset and evaluate its performance on a separate validation set.
- To deploy the trained model, making it accessible for stakeholders.

Exploratory Data Analysis

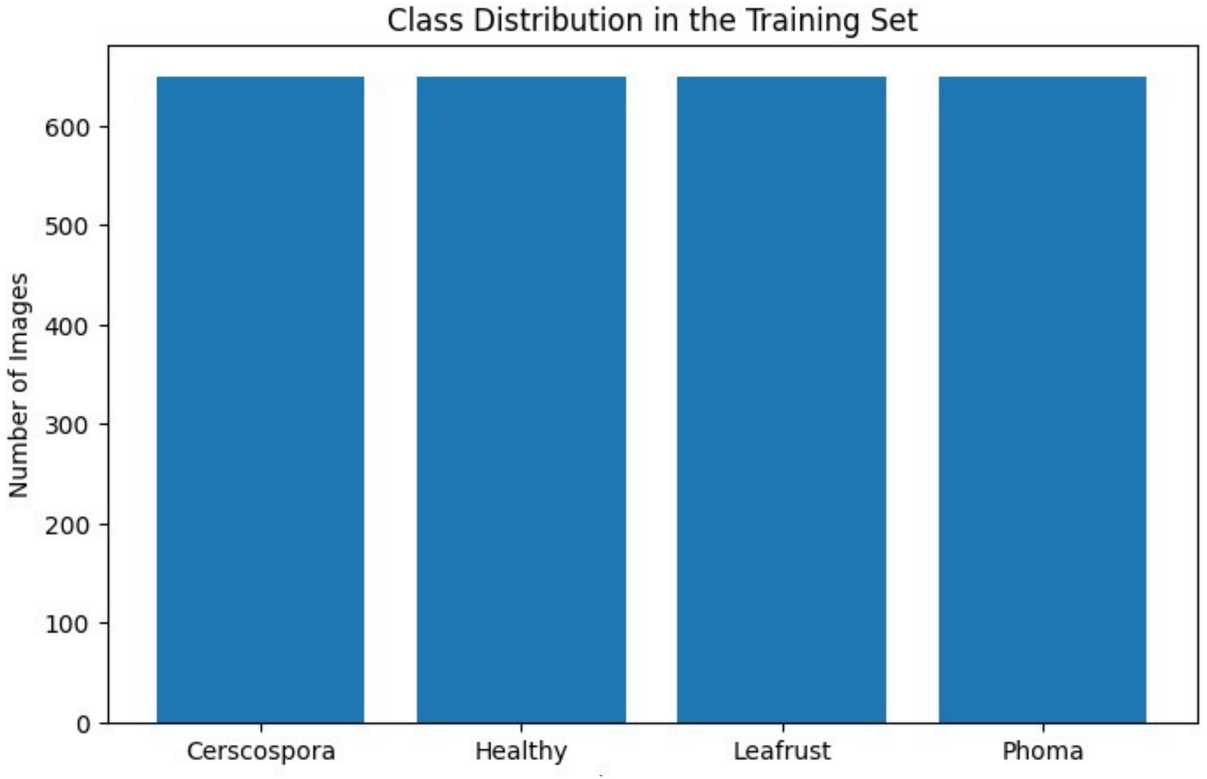


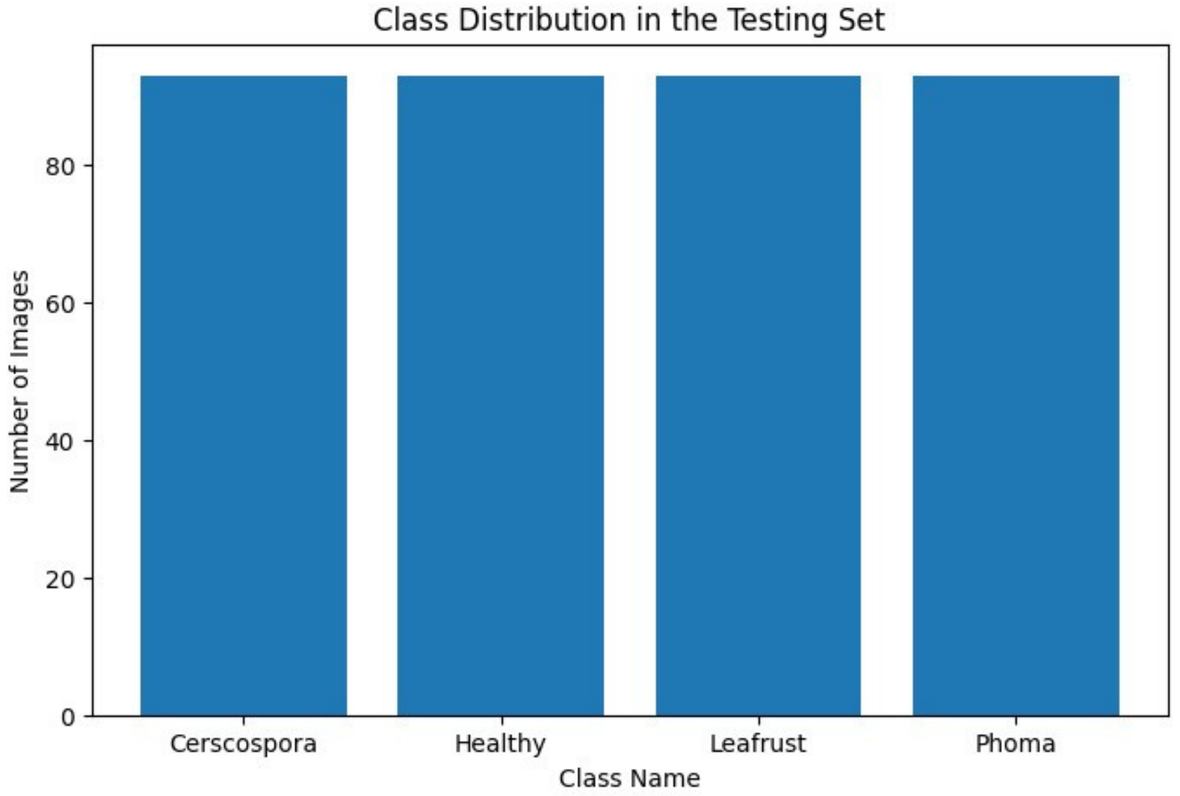
- We plot five images for each class as shown.

- From the plot, we can see the four classes of the Coffee Diseases i.e Cerscospora, Leaf-rust, Phoma and healthy leaves clearly labeled.

Exploratory Data Analysis

- We visualize class distributions in the train and test sets:
- The class distribution for the training set is 650 and for the testing set is 90 images.





MODELING

PARAMETERS	BASELINE MODEL	MODEL TWO	MODEL THREE	MODEL FOUR
LAYERS	2	2	3	4
MAX POOL	-	2	3	4
OPTIMIZER	ADAM	ADAM	ADAM	ADAM
ACCURACY	0.891	0.987	0.989	0.991

Conclusions

- Each model iteration showed significant improvement in accuracy, indicating the effectiveness of incorporating more sophisticated architectures.
- Convolutional layers proved crucial in capturing hierarchical and intricate features within leaf images, resulting in higher classification accuracy.
- Utilizing the Adam optimizer contributed to model convergence and enhanced performance across different architectures.

Recommendations

- Diversify image sources
- Dedicated computational resources

Next Steps

- Continuous Monitoring and Retraining.
- Education and Awareness Campaigns.
- Long-Term Sustainability Planning.

THANK YOU